**E1 244: Detection & Estimation Theory** **Jan-May 2017**

Lecture 7 — January 24

*Lecturer: Aditya Gopalan* *Scribe: B Phanidhar & Soumya Subhra Banerjee*

# Composite Hypothesis Testing

## 7.1 Recap of Composite Hypothesis Testing

The simple hypothesis-testing problem is called so because each of the two hypothesis corresponds to only a single distribution for the observation.On the contrary,in a composite hypothesis-testing problem there may be several distributions corresponding to the same hypothesis.The number of hypotheses however remains same ,i.e two for binary hypothesis testing.Here, we shall restrict ourselves to two hypotheses and exclude discussions on $M$-ary hypothesis testing.

The chief components to define composite hypothesis testing are as follows:

- **Parameter Space**: $\Lambda$
- **Observation Space**: $\Gamma$
- **Probability distributions**: $\mathbb{P}_\theta$ ,defined on $\Gamma, \forall\ \theta \in \Lambda$
  Note,$\Lambda$ can be a countable or an uncountable set.Broadly, some of the distributions pertaining to it's elements correspond to $H_0$ and some correspond to $H_1$.
- **Decision rule(non-randomized)**: $\delta : \Gamma \to \{0, 1\}$
- **Costs**: $C(i, \theta)$: Cost for declaring $H_i$ when the true distribution is $\mathbb{P}_\theta$
  $\forall$ i$\in\{0,1\}$ ,$\forall\theta \in \Lambda$ .We can have different costs modelled on different values of $\theta$.
- **Conditional Risks**: for a decision rule $\delta$, is defined for any given $\theta$ as,

  $$\mathbb{R}_\theta(\delta) := \mathbb{E}_\theta[C(\delta(Y), \theta)]$$

  where $\mathbb{E}_\theta$ denotes expectation assuming $Y \sim \mathbb{P}_\theta$.
- **Bayes Risk**: The average conditional risk under a prior $\pi$ on $\Lambda$ ,

  $$r(\delta) := \mathbb{E}[C(\delta(Y), \Theta)]$$
  $$\Theta \sim \pi \text{ on } \Lambda$$

  where $\mathbb{E}[.]$ denotes expectation assuming $\Theta \sim \pi$ on $\Lambda$.

*Note* 1. $\mathbb{E}[.]$ denotes expectation in the experiment which has two sources of randomness,

1. $\Theta \sim \pi$ ,over $\Lambda$
2. $Y \sim \mathbb{P}_\theta$ ,over $\Gamma$

This is in fact, a generalization of a simple hypothesis testing problem.

# 7.2 Finding the Bayes rule

The Bayes risk is given by,

$$\begin{aligned} r(\delta) &= \mathbb{E}[C(\delta(Y), \Theta)] \\ &= \mathbb{E}[\mathbb{E}[C(\delta(Y), \Theta)|Y]] \text{ (by iterated expectation)} \end{aligned}$$

Observe that if we fix $Y = y$, $\delta(y)$ can take values either 0 or 1.Hence, we can write $r(\delta)$ as follows,

$$r(\delta) = \int_\Gamma \begin{cases} \mathbb{E}[C(0,\Theta)|Y = y]p(y)dy, \text{ if } \delta(y) = 0 \\ \mathbb{E}[C(1,\Theta)|Y = y]p(y)dy, \text{ if } \delta(y) = 1 \end{cases} \tag{7.1}$$

To reduce $r(\delta)$ we choose $\delta(y)$ to take the minimum of the above expectations.
Since we have to design $\delta(y) \in \{0{,}1\}$ $\forall y \in \Gamma$, an optimal Bayes rule would be,

$$\delta(y) = \begin{cases} 1, \text{ if } \mathbb{E}[C(1,\Theta)|Y = y] < \mathbb{E}[C(0,\Theta)|Y = y] \\ 0, \text{ o.w.} \end{cases} \tag{7.2}$$

(Note: Breaking ties does not affect the Bayes risk.)
The above expectations are expected costs under the aposteriori distribution on $\Theta$ after observing $y$ , i.e., with no observation initial priors are used,the prior is updated with a posterior distribution on observing some $y$.

*Remark* 1. A Bayes rule with prior $\pi$ on $\Theta$, and observation $y \equiv$ A Bayes rule with prior : (posterior distribution of $\Theta|Y$), and observation $\phi$ (i.e., no observation)

## 7.2.1 Uniform cost (an important special case):

Uniform costs: Uniform costs over disjoint subsets of $\Lambda$ is defined as the following.

$$\begin{aligned} \Lambda = \Lambda_0 \cup \Lambda_1, \Lambda_0 \cap \Lambda_1 = \phi \\ \&, C(i,\theta) = C_{ij} \forall \theta \in \Lambda_j \\ \text{for } i, j \in \{0, 1\} \end{aligned}$$

| | $\theta \in \Lambda_0$ | $\theta \in \Lambda_1$ |
|---|---|---|
| i=0 | $C_{00}$ | $C_{01}$ |
| i=1 | $C_{10}$ | $C_{11}$ |

Figure 7.1: Uniform Cost Table over two disjoint subsets for Composite Hypothesis Testing.

In other words,the uniform cost of declaring a hypothesis $H_i$ is same for all $\theta$ belonging to a disjoint subset of $\Lambda$.
Assume $C_{11} < C_{01}$. Thus,under uniform costs,

$$
\begin{aligned}
\mathbb{E}[C(0,\Theta)|Y=y] &\gtreqless \mathbb{E}[C(1,\Theta)|Y=y] \\
\text{iff,}\quad C_{00}\mathbb{P}[\Theta \in \Lambda_0|Y=y] + C_{01}\mathbb{P}[\Theta \in \Lambda_1|Y=y] &\gtreqless C_{10}\mathbb{P}[\Theta \in \Lambda_0|Y=y] + C_{11}\mathbb{P}[\Theta \in \Lambda_1|Y=y] \\
\text{iff,}\quad \frac{\mathbb{P}[\Theta \in \Lambda_1|Y=y]}{\mathbb{P}[\Theta \in \Lambda_0|Y=y]} &\gtreqless \frac{(C_{00}-C_{10})}{(C_{11}-C_{01})} \\
\text{iff,}\quad \frac{\mathbb{P}[Y=y|\Theta \in \Lambda_1]\mathbb{P}[\Theta \in \Lambda_1]}{\mathbb{P}[Y=y|\Theta \in \Lambda_0]\mathbb{P}[\Theta \in \Lambda_0]} &\gtreqless \frac{(C_{00}-C_{10})}{(C_{11}-C_{01})}, \text{(by Bayes formula)} \\
\\
\text{iff,}\quad L(y) &\gtreqless \tau \\
\text{where,}\quad L(y) &:= \frac{\mathbb{P}[Y=y|\Theta \in \Lambda_1]}{\mathbb{P}[Y=y|\Theta \in \Lambda_0]}, \\
\tau &:= \frac{\pi_0(C_{00}-C_{10})}{\pi_1(C_{11}-C_{01})}, \\
\text{with,}\quad \pi_i &:= \mathbb{P}[\Theta \in \Lambda_i], i = 0, 1
\end{aligned}
$$

*Remark* 2. With uniform costs over $\Lambda_0$ and $\Lambda_1$, the optimal Bayes rule is again a likelihood ratio test.

*Remark* 3. We can write

$$\mathbb{P}[Y = y|\Theta \in \Lambda_1] = \int_{\Lambda} p_\theta(y) w_1(\theta) d\theta = \int_{\Lambda_1} p_\theta(y) w_1(\theta) d\theta$$

$$\text{where , } \quad w_1(\theta) = \begin{cases} 0, & \text{if } \theta \notin \Lambda_1 \\ \frac{\pi(\theta)}{\int_{\Lambda_1} \pi(\theta^{`}) d\theta^{`}}, & \text{if } \theta \in \Lambda_1 \end{cases}$$

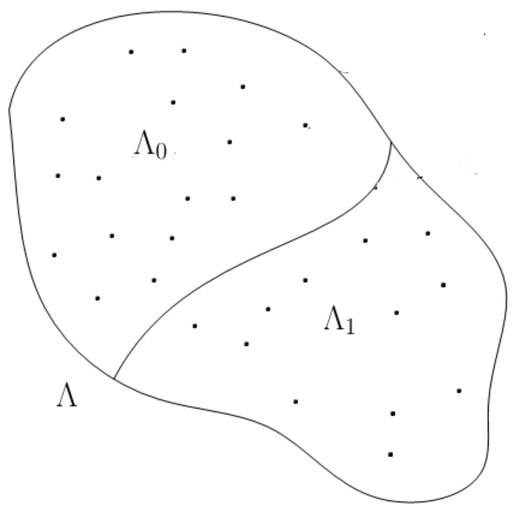


Figure 7.2: The Parameter space $\Lambda$ and with distribution $\Theta \sim \pi$, over $\Lambda$.

($\pi$ puts a distribution (i.e., $\pi(\theta), \theta \in \Lambda$) over whole $\Lambda$.We can say, $w_1(\theta)$ is the scaled probability of what $\pi$ puts on a $\theta$ in $\Lambda_1$.)

**Example 7.2.1.** 2-D Location testing with Gaussian noise

Here the null hypothesis has a mean zero bi-variate Gaussian distribution which denotes pure noise (denoted by the origin in figure 7.3), while the alternate hypothesis has multiple bi-variate Gaussian distributions with mean lying on the circle of radius $a$.

Goal: Detect signal (point on circle) or no signal (origin) in Gaussian noise.

$$\Gamma = \mathbb{R}^2, [i.e., Y = (Y_1, Y_2)^T], \sigma^2 \in \mathbb{R}_+, a \in \mathbb{R}_+,$$

$$H_0 : Y \sim \mathcal{N}(\left(\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right), \sigma^2 I)$$
$$H_1 : Y \sim \mathcal{N}(\left(\begin{smallmatrix} a\cos\psi \\ a\sin\psi \end{smallmatrix}\right), \sigma^2 I),$$
$$\text{with,} \Psi \sim unif[0, 2\pi],$$

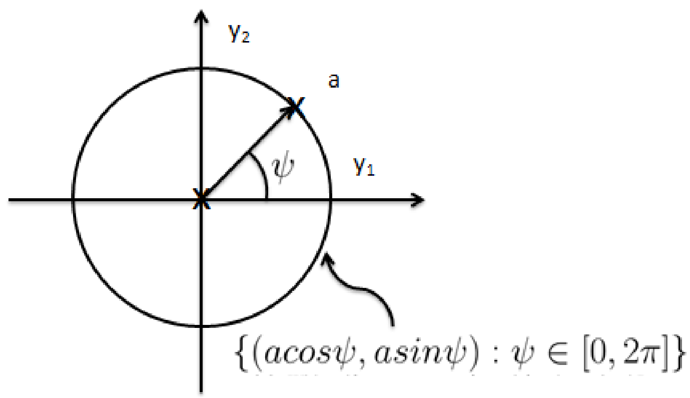


Figure 7.3: Distributions of hypotheses in 2-D location testing with Gaussian noise.

This problem is called "NON COHERENT DETECTION PROBLEM" in communication theory.
**Parameter space:**

$$\begin{aligned} \Lambda &= \overbrace{\{0,a\}}^{\text{“}\theta_1\text{”}} \times \overbrace{[0,2\pi]}^{\text{“}\theta_2\text{”}} \\ \Lambda_0 &= \{\theta \in \Lambda : \theta_1 = 0\} \\ \Lambda_1 &= \{\theta \in \Lambda : \theta_1 = a\} \\ \text{where,} \quad \theta &\equiv (\theta_1, \theta_2) \end{aligned}$$

Assume uniform costs over $\Lambda_0, \Lambda_1$.
(Note, $\theta_2$ pertains to $\psi$ in figure 7.3.Also, $\theta_1 = 0$,with any value of $\theta_2$, pertains to the point at origin.)
We can write,

$$L(y) = \frac{P(y|\Theta \in \Lambda_1)}{P(y|\Theta \in \Lambda_0)} \tag{7.3}$$

Where,
Denominator:

$$P(y|\Theta \in \Lambda_0) = \frac{\exp(\frac{-y_1^2 - y_2^2}{2\sigma^2})}{2\pi\sigma^2}$$

Numerator:

$$P(y|\Theta \in \Lambda_1) = \frac{1}{2\pi}\int_0^{2\pi} \frac{\exp(\frac{-(y_1-a\cos\theta_2)^2-(y_2-a\sin\theta_2)^2}{2\sigma^2})}{2\pi\sigma^2} d\theta_2$$

(by averaging over $\theta_2 \in [0, 2\pi]$)

hence,

$$L(y) = \frac{\exp(\frac{-a^2}{2\sigma^2})}{2\pi}\int_0^{2\pi} \exp(\frac{a(y_1\cos\theta_2 + y_2\sin\theta_2)}{\sigma^2})d\theta_2 \tag{7.4}$$

Changing to polar co-ordinates,

$$r := \sqrt(y_1^2 + y_2^2)$$
$$\phi := \arctan(\frac{y_2}{y_1})$$

We get $L(y)$ as follows,

$$\begin{aligned} L(y) &= \frac{\exp(\frac{-a^2}{2\sigma^2})}{2\pi}\int_0^{2\pi}\exp(\frac{ar\cos(\theta_2-\phi)}{\sigma^2})d\theta_2 \\ &= \frac{\exp(\frac{-a^2}{2\sigma^2})}{2\pi}\int_0^{2\pi}\exp(\frac{ar\cos u}{\sigma^2})du \text{ ,where } u = \theta_2 - \phi. \\ &= \exp(\frac{-a^2}{2\sigma^2}).I_0(\frac{ar}{\sigma^2}) \end{aligned}$$

Here $I_0(x)$ is the modified Bessel function,given by,

$$I_0(x) := \frac{1}{2\pi}\int_0^{2\pi}\exp(x\cos u)du$$

Note that $L(y)$ does not depend on $\theta_2$ in the final expression.Hence,

$$L((\begin{smallmatrix} y_1 \\ y_2 \end{smallmatrix})) \gtreqless \tau$$
$$\Leftrightarrow r = \sqrt(y_1^2 + y_2^2) \gtreqless I_0^{-1}(\tau\exp(\frac{a^2}{2\sigma^2}))\frac{\sigma^2}{a}$$

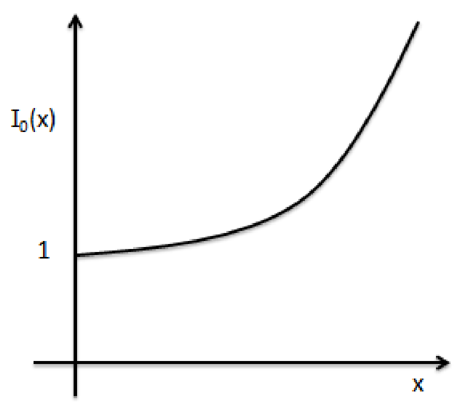


Figure 7.4: The modified Bessel function.